

Contents

Foreword to the First English Edition	xiii
Preface	xv
Acknowledgements	xix

Part I

Introduction

1

Lecture 1

3

Main functions of proteins. Amino acid sequence determines the three-dimensional structure, and the structure determines the function. The reverse is not true. Fibrous, membrane, and globular proteins. Primary, secondary, tertiary, and quaternary structures of proteins. Domains. Co-factors. Active sites and protein globules. Protein biosynthesis; protein folding in vivo and in vitro. Post-translational modifications.

Recommended additional reading 12

Part II

Elementary Interactions in and Around Proteins

15

Lecture 2

17

Amino acid residues in proteins. Backbone and side chains. Stereochemistry of natural L-amino acids. Peptide bonds. Covalent interactions and quantum mechanics. Heisenberg's principle of uncertainty. Covalent bonds and angles between them. Their vibrations. Dihedral angles. Rotation around the covalent bonds. Potential barriers for rotations. Peptide group: why is it flat and rigid? *Trans*- and *cis*-prolines.

References 25

Lecture 3

27

Quantum mechanics, Pauli Exclusion Principle, and non-covalent interactions. Van der Waals interaction: attraction at long distances, repulsion at short distances. Potential of the van der Waals interaction. Typical radii of atoms. Why *cis*-conformations of peptide bonds are rare. Allowed conformations of amino acid residues. Ramachandran plots for glycine, alanine, valine, proline.

References 37

Lecture 4 39

Influence of water environment. Hydrogen bonds. Their electrostatic origin. Their energy. Their geometry in crystals. Ice. Hydrogen bonds in water are distorted. Notes on entropy and free energy. Hydrogen bonds in the protein chain replace the same bonds of this chain to water; as a result, the hydrogen bonds in proteins—in an aqueous environment—are entropy-driven bonds.

References 50

Lecture 5 51

Water, hydrophobicity, and proteins. Boltzmann distribution. Elements of thermodynamics. Free energy and chemical potential. Hydrophobic interactions: entropic forces. Their connection with the necessity to saturate hydrogen bonds in water. Origin of hydrophobic interactions. The strength of hydrophobic interactions depends on temperature. Hydrophobic effect is responsible for formation of compact globules. Water-accessible surface of amino acids and their hydrophobicity.

References 65

Lecture 6 67

Influence of an aqueous environment on electrostatic interactions. Electric field near the surface and inside the protein globule. Permittivity. Electrostatic interactions and corpuscular structure of the medium. Debye-Hückel effect: shielding of charges in saline solutions. Measuring of electric fields in proteins by protein engineering. Electrostatics in water is entropic. Disulfide bonds. Coordinate bonds. Entropic forces—again.

References 81

Part III
Secondary Structures of Polypeptide Chains 83**Lecture 7** 85

Secondary structure of polypeptides. Chirality of helices. Helices: 2_7 , 3_{10} , α , π , poly(Pro)II. The main secondary structures: hydrogen bonds and Ramachandran maps of allowed and disallowed conformations of amino acid residues. Antiparallel and parallel β -structure. Small secondary structures: β -turns, β -bulges. What is the “coil”? Methods of experimental identification of secondary structure.

References 99

Lecture 8 101

Elements of statistical mechanics. Temperature; its connection with the change of entropy with the energy increase. Probability of states with different

energy (Boltzmann-Gibbs distribution). Partition function and its connection with the free energy. Conformational changes. The first-order ("all-or-none" in small systems) phase transitions, the second-order phase transitions, and non-phase transitions. Kinetics of free-energy barrier overcoming during "all-or-none" conformational changes: what it looks like at the level of a molecule and at the level of an ensemble of molecules. The transition-state theory of reaction rates. Parallel and sequential processes. Typical times of diffusion processes. The mean free path of molecules in water.

References 121

Lecture 9 123

Free energy of initiation and elongation of α -helices in a homopolypeptide. Landau's theorem and the non-phase character of the helix-coil transition. The size of the cooperative region at the helix-coil transition. α -Helix stability in water. α -Helix and β -sheet boundaries. β -Structure stability in water. Rate of α -helix, β -hairpin, and β -sheet formation.

References 136

Lecture 10 139

20+2+1 gene-coded amino acids in proteins. Properties of amino acid side chains. Amino acids in secondary structures. Alanine, glycine, proline, valine. Branched side chains. Non-polar, short polar, and long polar side chains. Capping of α -helices. Charged side chains; pH dependence of their charges. Hydrophobic surfaces on protein secondary structures.

References 146

Part IV Protein Structures 149

Lecture 11 151

Fibrous proteins, their functions, their regular primary and secondary structures; α -keratin, silk β -fibroin, collagen. Packing of secondary structures. α -Helical coiled coils. Collagen triple helix. Assisted folding of collagen. Packing of long α -helices and large β -sheets. Matrix-forming proteins; elastin. Genetic defects of protein structures and diseases. Amyloids. Kinetics of their formation.

References 161

Lecture 12 165

Membrane proteins; peculiarities of their structure and function. Transmembrane α -helices, transmembrane β -barrels. High cost of hydrogen bonds in a lipid environment. Bacteriorhodopsin; receptors and G-proteins; porin;

photosynthetic reaction center. Transmitting channels. Selective permeability of membrane pores. The photosynthetic reaction center at work. Tunneling. Electron-conformational interaction. Assisted and spontaneous folding of membrane proteins.

References 179

Lecture 13 181

Water-soluble globular proteins. Is the protein structure the same both in crystals and in solution? Simplified presentation of protein globules. Structural classes, architectures, topologies, folding patterns. Structure of β -proteins: β -sheets, their aligned and orthogonal packing. Meanders, Greek keys, jellyrolls, blades, prisms. β -Structure in β -proteins is mainly antiparallel. Topology of β -proteins.

References 196

Lecture 14 199

Structure of α -proteins. Bundles and layers of helices. Model of the quasi-spherical α -helical globule. Close packing of α -helices. Structure of α/β -proteins: parallel β -sheet covered with α -helices, α/β -barrel. Topology of β - α - β subunits. Structure of α + β proteins. The absence of direct connection between overall protein architecture and its function, though the active site position is often determined by the overall protein architecture.

References 212

Lecture 15 215

Classification of protein folds. The absence of observable "macroevolution" of protein folds and the presence of observable "microevolution" of their detailed structures. Gene duplication and protein specialization. Evolution by reconnection of domains. "Standard" protein folds. Typicality of "quasi-random" patterns of amino acid sequences in primary structures of globular proteins in contrast to periodic sequences of fibrous proteins and blocks in the sequences of membrane proteins. Physical principles of architectures of protein globules. The main features observed in protein globules: separate α - and β -layers; rare over-crossing of loops; rare parallel contacts of secondary structures adjacent in the chain; rare left-handedness of β - α - β superhelices. "Energy-" and "entropy-defects" of rarely observed structures and connection between these "defects" and rarity of the amino acid sequences capable of stabilizing the "defective" structures. Natural selection of protein folds. "Multitude principle."

References 229

Lecture 16 233

What secondary structure can be expected for random and quasi-random amino acid sequences? Domain construction of the folds formed by long quasi-random sequences. Quasi-Boltzmann statistics of the elements of protein structures. These statistics originate from the physical selection of stable protein structures. Influence of element stability on selection of sequences supporting the fold of a globular protein; or, why some protein structures occur frequently while others are rare. What structure— α or β —should be usually expected in the center of a large globule? Connection between “entropy-defects” and “energy-defects.” Do globular proteins emerge as “selected” random polypeptides? Selection of “protein-like” sequences in protein engineering experiments.

References 249

Part V

Cooperative Transitions in Protein Molecules 251

Lecture 17 253

“Well-folded” and “natively disordered” (or “intrinsically disordered”) proteins. Protein denaturation. Cooperative transitions. Reversibility of protein denaturation. Denaturation of globular protein is a cooperative “all-or-none” transition. Van’t Hoff criterion for the “all-or-none” transition. Heat and cold denaturation. Phase diagram for states of a protein molecule. What does denatured protein look like? The coil and the molten globule. Large-scale inhomogeneity of some “natively disordered” proteins. The absence of “all-or-none” phase transition during swelling of “normal” polymer globules.

References 271

Lecture 18 275

Denaturation of globular protein: why is it an “all-or-none” transition? Decay of closely packed protein core and liberation of side chains. Penetration of solvent into denatured protein; decay of the molten globule, subsequent gradual unfolding of the protein chain with increasing solvent strength. Energy gap between the native protein fold and all other globular folds of the chain. The main physical difference between a protein chain and random heteropolymers. The difference in melting between “selected” chains (with energy gap) and random heteropolymers.

References 286

Lecture 19 289

Protein folding in vivo and in vitro. Co-translational folding. Auxiliary mechanisms for in vivo folding: chaperones, etc. Spontaneous folding is

possible in vitro. Aggregation, the main obstacle to in vitro protein folding, and crowding effects in vivo. The “Levinthal paradox.” Protein folding experiments in cell-free systems; on various understandings of the term “in vitro.” Stepwise mechanism of protein folding. Discovery of metastable (accumulating) folding intermediates for many proteins. The molten globule is a common (but not obligatory) intermediate in protein folding under “native” conditions. The simplest (“two-state”) folding of some proteins proceeds without any accumulating intermediates. Folding nucleus.

References 304

Lecture 20 307

Two-state folding of small proteins: kinetic analog of the thermodynamic “all-or-none” transition. Two- and multi-state folding. Theory of transition states. Experimental identification and investigation of unstable transition states in protein folding. Φ -value analysis. Folding nucleus. Its experimental discovery by protein engineering methods. Nucleation mechanism of protein folding. Native and non-native interactions in the nucleus. Folding nucleus is less specific and less “invariant” than the native protein structure.

References 320

Lecture 21 323

Solution of “Levinthal paradox”: a set of fast folding pathways (a “folding funnel” *with* phase separation) automatically leads to the most stable structure. It is necessary only to have a sufficient energy gap separating the most stable fold from others. Volume of conformational space at the level of secondary structure formation and assembly. Discussion of very slow folding of stable structure in some proteins: serpins. “Chameleon” proteins. Misfolding. Notes on “energy funnels” and “free-energy landscapes” of the folding protein chains. Consideration of the unfolding and folding sides of the free-energy barrier. The detailed balance law. Protein structures: physics of folding and natural selection of chains capable of folding.

References 344

Part VI

Prediction and Design of Protein Structure 347

Lecture 22 349

Protein structure prediction from amino acid sequences. Specific amino acid sequences of globular, membrane, fibrous, and intrinsically disordered proteins. Recognition of protein structures and their functions using homology of sequences. Key regions and functional sites in protein structures. Profiles for primary structures of protein families and multiple alignments. Detection

of stable elements of protein structures. “Templates” of elements of protein structures. Predicted structures are unavoidably judged from only a part of interactions occurring in the chain. As a result, we have probabilistic predictions. Interactions that stabilize and destroy secondary structures of polypeptide chains. Calculation of secondary structures of non-globular polypeptides. Prediction of protein secondary structures.

References 364

Lecture 23 367

Overview of approaches to prediction and recognition of tertiary structures of proteins from their amino acid sequences. Protein fold libraries. Recognition of protein folds by threading. Prediction of common fold for a set of remote homologs reduces uncertainty in fold recognition. Structural genomics and proteomics. Bioinformatics. Advances in modeling of protein folding. Protein engineering and design. Almost identical sequences can produce different folds with different functions. Advances in design of protein folds.

References 381

Part VII 385

Physical Background of Protein Functions

Lecture 24 387

Protein function and protein structure. Elementary functions. Binding proteins: natively disordered proteins, DNA-binding proteins, immunoglobulins. Enzymes. The active site as a “defect” of globular protein structure. Protein rigidity is crucial for elementary enzyme function. Catalytic and substrate-binding sites. Inhibitors. Mechanism of enzymatic catalysis; activation of enzymes. Example: serine proteases. Transition state theory and its confirmation by protein engineering. Abzymes. Specificity of catalysis. “Key-lock” recognition.

References 406

Lecture 25 409

Combination of elementary functions. Transition of substrate from one active site to another. “Double sieve” increases specificity of function. Relative independence of protein folds from their elementary catalytic functions. Different catalytic sites can perform the same job. *Ser-proteases* and *metalloproteinases*. Multicharged ions. Visible connection between protein fold and protein environment. Combination of elementary protein functions and flexibility of protein structure. Induced fit. Mobility of protein domains. Shuffling of domains and protein evolution. Domain structure: kinases, dehydrogenases. Co-factors. Allostery: interaction of active sites.

Allosteric regulation of protein function. Allostery and protein quaternary structure. Hemoglobin and myoglobin. Mechanochemical cycle. Kinesin: a bipedal protein. Mechanism of muscle contraction. Rotary motors.

References 427

Appendices 429

Appendix A

Theory of globule-coil transitions in homopolymers 431

References 435

Appendix B

Theory of helix-coil transitions in homopolymers 436

References 439

Appendix C

Statistical physics of one-dimensional systems and dynamic programming 440

References 446

Appendix D

Random energy model and energy gap in the random energy model 448

References 452

Appendix E

How to use stereo drawings 453

Problems with solutions and comments 457

Index 501